Jl. of Technology and Teacher Education [illegible]

"Unnatural How Natural It Was": Using a Performance Task and Simulated Classroom for Preservice Secondary Teachers to Practice Engaging Student Avatars in Scientific Argumentation

JAMIE N. MIKESKA
ETS, USA
jmikeska@ets.org

CALLI SHEKELL
Thiel College, USA
cshekell@thiel.edu

JENNIFER DIX
Towson University, USA
jdix4@students.towson.edu

PAMELA S. LOTTERO-PERDUE
Towson University, USA
plottero@towson.edu

Facilitating discussions is a key approach that science teachers use to engage students in scientific argumentation. However, learning how to facilitate argumentation-focused discussions is an ambitious teaching practice that can be difficult to learn how to do well, especially for preservice teachers (PSTs) who typically have limited opportunities to tryout and refine this teaching practice. This study examines secondary PSTs' perceptions and engagement with a science performance task—used within an online, simulated classroom consisting of five middle school student avatars—to practice this ambitious teaching practice. Findings showed that the PSTs had a strong understanding of the discussion's primary goal and perceived the task components to be easy to understand, useful in helping them prepare for the simulated discussion, and an authentic representation of what middle school students would say and do. In addition, while the PSTs attended to similar content and pedagogical features within their facilitated discussion, they varied in their ability to successfully facilitate the discussion. This study adds to the growing literature on innovative, technology-based solutions for supporting teacher learning and points to one productive approach that can be incorporated within secondary science teacher education as an approximation of practice of this ambitious teaching practice.

Engaging students in argument from evidence is one of the science and engineering practices identified in the *Next Generation Science Standards* (NGSS) as important for student learning (NGSS Lead States, 2013). Scientific argumentation involves students in generating and defending scientific claims using evidence-based reasoning, as well as comparing and critiquing one another's ideas as students work to persuade one another and build consensus (Berland & Reiser, 2009; Osborne et al., 2016). Facilitating discussions is a key approach that science teachers can use to engage students in this practice (Cartier et al., 2013; Simon & Richardson, 2009). Yet, learning how to facilitate discussions that engage K-12 students in productive scientific argumentation is an ambitious teaching practice that is difficult to learn how to do well and one that preservice teachers (PSTs) tend to have limited opportunities to tryout and refine.

Research findings have illustrated both the successes and challenges of engaging students productively in scientific argumentation and have identified many factors that can impact teachers' abilities to engage their students in scientific argumentation, such as teachers' knowledge about argumentation, their access to high-quality curriculum resources, and their perspectives about its importance (Colley & Windschitl, 2016; McNeill et al., 2016, 2017; Osborne et al., 2013; Sadler, 2006). Within the last decade, additional research has explored various approaches that can be used to help teachers learn how to facilitate scientific argumentation (Marco-Bujosa et al., 2017; McNeill et al., 2016, 2017; Osborne et al., 2013; Sadler, 2006). Yet, findings have shown mixed results, suggesting that work to expand the available approaches would be an important contribution to science teacher education.

This study's primary focus was to examine how one performance task coupled with Mursion's® online, simulated classroom consisting of five middle school student avatars (Figure 1) could be used to provide opportunities for PSTs to practice one ambitious science teaching practice: facilitating argumentation-focused discussions. In this study, we investigated the PSTs' perceptions and use of this innovative tool to engage them in an approximation of practice. The study addresses the following research questions (RQs):

- RQ1: How do the PSTs perceive the clarity, authenticity, usefulness, and discussion goal of the performance task?
- RQ2: How well and in what ways do the PSTs facilitate argumentation-focused science discussions within the simulated classroom?

The first research question provides empirical evidence to determine whether the PSTs understand the intent and purpose of the performance task and to understand the extent to which they may value its potential use as a tool within teacher education settings. The second research question provides insight into the PSTs' actual engagement with the performance task, which helps to illuminate the ways in which such a task could be used to provide specific learning opportunities to PSTs. Collectively, these findings can provide answers about the potential and feasibility of such tools to productively complement other instructional approaches when integrated within PST preparation programs.

Figure 1. Mursion's® Middle School Classroom. Image courtesy of Mursion, Inc.

We begin with a brief description of the theoretical framework undergirding this study – a practice-based theory of teacher learning. We then provide background into research that examines: (1) students' engagement in and teachers' facilitation of scientific argumentation and (2) the use of technologically mediated simulations to support teacher learning. After that, we move onto a discussion of the study's context, methods, and data analysis approach. We end by sharing findings addressing each research question and implications for how such innovative technology tools could be leveraged to support teacher learning within teacher education contexts.

THEORETICAL FRAMEWORK: PRACTICE-BASED THEORY OF TEACHER LEARNING

Practice-based teacher education has been lauded as one solution to address the widespread challenge of providing substantive, frequent, and meaningful opportunities for PSTs to rehearse key aspects of complex instructional practice (Ball & Forzani, 2009; Forzani, 2014; Francis et al., 2018; Grossman, Compton, et al., 2009; Lampert, 2009; Zeichner, 2012). This theory argues that teacher learning is directly tied to opportunities for them to learn in and from their practice, with recent research illustrating that PSTs are more effective when their preparation provides opportunities for such practice (Francis et al., 2018; Goodson et al., 2019). As noted by Grossman, Compton, et al. (2009), these learning opportunities can include a range of different pedagogies of practice, such as engaging PSTs in representing practice through sharing written cases, decomposing practice through analyzing video examples, or approximating practice in situations of reduced complexity.

Approximations of practice involve PSTs in trying out aspects of the work of teaching, such as practicing interpreting and eliciting student ideas or facilitating discussions, albeit in situations where they do not have to contend with the full complexity of instructional challenges (Grossman, Hammerness, et al., 2009). In science education, face-to-face rehearsals have been one of the key approaches used to engage PSTs in approximations of practice. In these rehearsals, the role of the "student" is played by one or more adults – typically other PSTs, the teacher educator, or other trained adults – as the PST tries out a novel teaching practice (Arias & Davis, 2019; Benedict-Chambers, 2016; Benedict-Chambers & Aram, 2017; Benedict-Chambers et al., 2020; Davis et al., 2017; Masters, 2020).

More recently, the field has seen an increase in the use of digital practice spaces, including virtual classroom environments, to engage PSTs in rehearsals within science education (Bell, 2019; Lottero-Perdue et al., 2020; Mikeska & Howell, 2020; Straub et al., 2015). Yet, questions remain about how these types of innovative technology-based tools are perceived and if and how they can be used by PSTs to support their learning. This study explores how a performance task used within an online simulated classroom environment could be used to engage PSTs in approximating one ambitious science teaching practice: facilitating discussions that engage students in scientific argumentation.

BACKGROUND

Learning How to Engage Students in Scientific Argumentation

Research has suggested that engaging students in scientific argumentation is important to support student learning. Opportunities for students to practice engaging in argumentative discourse has been linked to increased student engagement, understanding of disciplinary concepts and practices, and critical thinking and decision-making (Kuhn, 2010; McNeill et al., 2016; Osborne et al., 2013). Here, we define scientific argumentation as a process that involves students in two complementary aspects – argument construction and argument critique (Mikeska & Howell, 2020). Argument construction involves students in generating, defending, and refining scientific claims and using evidence-based reasoning to support and refute such claims. Argument critique involves students in comparing and critiquing scientific claims and using evidence-based reasoning to persuade one another. Both aspects of argumentation are in service of coming to consensus about scientific explanations. Collectively these two aspects attend to the structural and dialogic components of argumentation that have been highlighted in the science educational literature (Gonzalez-Howard & McNeill, 2019; Grooms et al., 2018; Jimenez-Aleixandre & Erduran, 2008).

Within the last two decades, several research studies have examined how K-12 students engage in argumentative discourse in the context of science instruction and the varied factors that relate to the productive nature of this engagement (Colley & Windschitl, 2016; McNeill, 2011; McNeill et al., 2016, 2017). For example, studies have shown that K-12 students are capable of successfully engaging in key aspects of productive scientific argumentation including generating scientific claims, providing and probing for evidence-based reasoning to support scientific claims, critiquing others' claims and evidence-based reasoning, and offering rebuttals and counter arguments. Studies have also suggested that opportunities students have to engage in scientific argumentation can be impacted by their own and their teachers' understanding of the key characteristics of scientific argumentation and beliefs about its importance; the curriculum materials that teachers use to guide their science instruction; and the ways in which teachers frame the learning goals for their students during science discussions (Berland & Hammer, 2012; Katsh-Singer et al., 2016; McNeill & Krajcik, 2008; McNeill & Pimentel, 2010; Sampson & Blanchard, 2012). Another consistent finding across studies is that scientific argumentation is supported by having a supportive classroom setting where risk-taking and differences in perspectives are valued and encouraged (Berland & Reiser, 2011; Henderson et al., 2018).

While there are varied approaches to supporting students to engage in productive scientific argumentation, one approach that has been used widely across many informal and formal instructional contexts is the use of student-centered discussions (Cartier et al., 2013; National Research Council, 2011). Research has suggested that teachers need access to scaffolded and comprehensive supports, typically through professional development, to learn how to support students in productively engaging in scientific argumentation. For example, one recent study (Fishman et al., 2017; Osborne et al., 2019) developed and implemented a multi-year professional development program with in-service elementary teachers to support them in learning how to facilitate student-led argumentative discourse in science. Findings from their study indicated that teachers improved in their ability to promote student argumentative discourse and their students also showed an uptake in engaging in specific discourse moves. In another line of research (Marco-Bujosa et al., 2017; McNeill et al., 2016, 2017), a team of researchers and science teacher educators developed a suite of online resources, including video exemplars, lesson planning tools, and learning modules, to support teachers in learning how to provide their students with opportunities to engage in productive scientific argumentation. Study findings were also positive in nature with noted improvements to teachers' self-efficacy, views about their students' capabilities, and attention to student learning goals during lesson planning after using these multimedia resources.

However, other studies provide evidence of less positive outcomes, illustrating how this teaching practice can be difficult for teachers to learn how to do well. For example, in one study (Osborne et al., 2013) researchers engaged two lead teachers in learning about how to use various scientific argumentation activities in the classroom and then had those lead teachers work to provide professional learning experiences to other teachers on how to integrate these activities into their own classrooms. The participating teachers did not consistently show improvement in their understanding of the science content or of argumentation instruction. These results suggest that exploring new tools that could be used to support teacher learning of this ambitious teaching practice, especially tools that could be used to expedite and transform such learning experiences, would be a useful contribution to the field.

Using Technologically Mediated Simulations in Teacher Education

One way to enable PSTs to practice facilitating argumentation discussions is by using simulations. We use the following definition of simulations within teacher education from a synthesis of recent work by Mikeska et al. (2021):

> Simulations are responsive learning spaces where preservice and inservice teachers can rehearse critical instructional practices or specific skills essential to the work of teaching in situations of reduced complexity. These learning spaces can target the interactive, in-the-moment, responsive work of teaching, such as eliciting student ideas or facilitating student-led discussion or the noninteractive components, such as planning, grading, providing written feedback on work, or interpreting student data. Simulations do not involve interactions with real students. Instead, they typically involve synchronous and human-driven interactions, where the participant interacts via a face-to-face format or through a technologically mediated environment with one or more adults who act as K-12 students. (p. 800)

In this study, we focus specifically on simulations that are technologically mediated, as opposed to those that involve, for example, peer-to-peer role play (e.g., Benedict Chambers, 2016).

There are multiple types of technologically mediated (or digital) simulations available for use in teacher education. Examples include Quest2Teach, Teacher Moments, Eliciting Learner Knowledge (ELK), simSchool, TeachLivE™, and Mursion® (Arici et al., 2016; Bondie et al., 2021; Christensen et al., 2011; Deale & Pastore, 2014; Thompson et al., 2019; Wang et al., 2021). It is beyond the scope of this paper to describe each of these. However, one important way to differentiate among these simulations is by whether they involve a teacher interacting with technology that is supported by a real-time "human in the loop" operating "behind" that interface. In what follows, we describe those that include a human in the loop since it is this type of simulation used in the present study.

These simulations are often referred to as mixed-reality simulations that utilize both a technological system and include one or more humans to operate. ELK, for example, is “a role-playing system that offers virtual sessions in which players can learn and practice discourse strategies on eliciting knowledge from conversational partners” (Wang et al., 2021, p. 2). One player plays the role of a student while the other a teacher. In this way, there are two humans in the loop, both of whom communicate via the online system interface by texting one another questions (teacher) and answers (student); thus, the players need not be in the same location to interact with one another. The goal of the teacher is to elicit ideas from the student; those “ideas” are included in the scenario that is shown to the player in the student role only. The teacher also receives information about what they are aiming to elicit. In a recent study, ELK was shown to have modest benefits with respect to the types of effective questioning strategies that PSTs used to elicit learner knowledge (Wang et al., 2019).

TeachLivE™ and Mursion® are simulations that involve a system in which avatars (which can be students, parents, other teachers, etc.) interact with the teacher. The avatars are simultaneously supported by artificial intelligence and must be operated by a professional simulation specialist—also called an interactor or human puppeteer—who is a highly trained human in the loop who plays the roles of the avatar(s) in the simulation (Dieker et al., 2014). The real-time interaction of the simulation specialist contributes to the feeling of authenticity within the simulated learning experience. Further, this system “[combines] the engaging features of face-to-face communication and the anonymity of online environments” (Straub, 2018, p. 2). Two studies suggest learning gains by teachers using TeachLivE in professional learning that can also extend to the classroom (Straub et al., 2014, 2015).

Using simulations like TeachLivE and Mursion® requires careful development of the scenario or task for the teacher *and* the development of training materials and processes to train the simulation specialist (Bondie & Dede, 2021). Additionally, the full benefit of using these simulations is achieved when teachers are prepared to go into the simulation, can use the simulation multiple times, and receive personalized coaching during and/or feedback after the simulation (Bondie & Dede, 2021; Mikeska et al., 2021). Allowing teachers to code transcripts of their interactions in the simulations has also shown promise in helping them change their questioning strategies (Lottero-Perdue et al., 2022; Wang et al., 2021). Coaching (e.g., by a teacher educator) during engagement in mixed reality simulations has also shown positive effects on teacher learning (Cohen et al., 2020).

What is practiced within simulations ranges across studies, yet these simulations invariably attend to interactive and communicative practices between teachers and students (or parents, etc.) and among students as facilitated by teachers. Simulations also vary with respect to speci c disciplinary contexts. Relevant to the present study is the work by Mikeska and colleagues (Mikeska et al., 2021; Mikeska & Howell, 2020) in which the Mursion® simulated classroom was used to help elementary teachers learn to facilitate argumentation discussions in science. Related work has examined the context of argumentation discussions in elementary mathematics (Howell et al., 2021) and engineering (Lottero-Perdue et al., 2020). To date, this work has been conducted at the elementary level. This study was designed to build out to the secondary level and consider whether a similarly designed discussion task deployed within a simulated classroom held similar promise as a practice space for PSTs studying to become middle school science teachers.

CONTEXT

Science Performance Task

The Keep It Cold science performance task involves two components: (1) a PST-facing written document describing the student learning goal for the discussion and information about where this discussion ts into a larger instructional sequence; and (2) a set of training materials for the simulation specialist who acts as all ve middle school student avatars during the discussion. In the Keep It Cold science task, PSTs facilitate a discussion between two groups of students, the goal of which is to come to consensus on a model describing heat transfer between the warm air and two separate cups of cold water that are made of different materials. The task materials explain that, before the Keep It Cold investigation, the student avatars completed a series of three other activities exploring heat transfer and the ow of energy (e.g., observing how different cup materials affected the rate of cooling for a cup of hot chocolate). Immediately prior to the PST-led discussion, student avatars in two small groups completed the Keep It Cold investigation where they observed and recorded the temperature of cold water in two different cups (foam vs. paper) every 10 minutes. After a half hour elapsed, the student groups drew a model to explain their observations about the differences in heat transfer between the air and the water in the cups. Each PST uses these student generated models and explanations, as well as the students' previous class activities, to facilitate a discussion between the students about how to best represent heat transfer in a class consensus model.

One of the groups, that of Savannah, Dev, and Ava, created a model that used arrows to show how heat energy from the warmer surrounding air moved towards the cups (Figure 2). For the paper cup, they showed the arrows moving into the cup, explaining that "heat energy gets into the paper cup." However, the arrows did not similarly move through the foam cup into the water inside of the foam cup. The group explained that "the foam stops the heat, so the temperature of the water does not increase like the temperature of the water in the paper cup." They also explained that the "heat energy ... can't get into the foam cup." In response to this idea from Savannah, Dev and Ava, one PST asked them to reflect back to the hot chocolate activity asking, "Did you feel any heat at all coming from the foam cup?" to prompt them to consider whether the foam completely blocked heat from entering (in the Keep It Cold investigation) or escaping (in the hot chocolate activity).

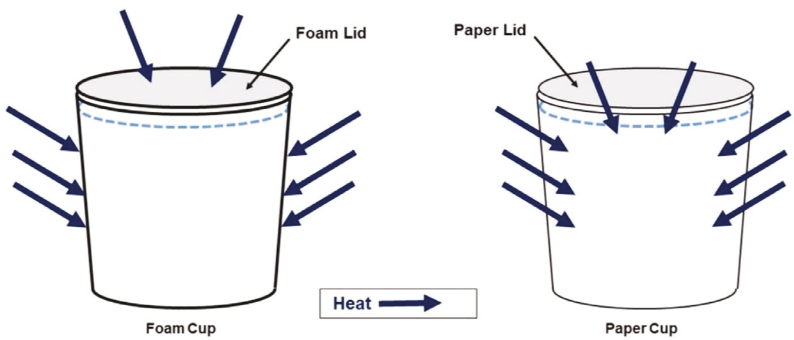

Figure 2. Savannah, Dev, and Ava's Model in the Keep It Cold Science Performance Task.

Savannah, Dev, and Ava's model correctly indicated the direction of heat transfer being from the warmer air to the cooler water. Their model could be improved by showing heat transfer arrows penetrating the foam wall since the cold water in the foam cup also increased in temperature over time. During the discussion, Jasmine and Ethan's attention to the data table may be instrumental in convincing Savannah, Dev, and Ava of this idea. Savannah, Dev, and Ava might also benefit from applying a small particle model to show the difference in temperature between the water in the foam and paper cups; this was a good suggestion from Jasmine and Ethan's critique of Savannah, Dev, and Ava's model. For example, one PST attempted to call Savannah, Dev and Ava's attention to the data table by asking, "Is there any evidence or any data that can support your model?"

Jasmine and Ethan's model used particles instead of arrows (Figure 3). They explained that in their model they "used different colors to show the different temperatures of particles" and wrote that the "cold particles" moved more slowly than the "warm particles." They also explained that more cold particles escaped from the paper cup than from the foam cup, stating that "the cold leaks out of the paper cups faster than the foam cups" and the "foam keeps most of the cold particles inside."

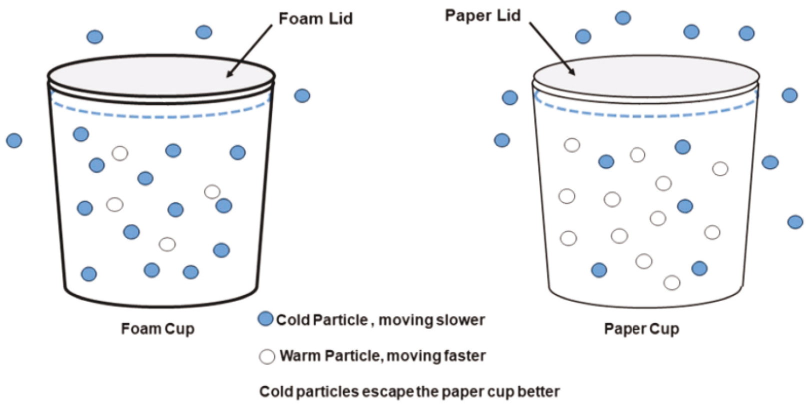

Figure 3. Jasmine and Ethan's Model in the Keep It Cold Science Performance Task.

Strengths of Jasmine and Ethan's model and explanation include that they draw from evidence in the data table of temperature over time and they attempt to represent different particle motion. Improvements would focus on helping them understand that there are only water particles (not cold or warm water particles), and that the direction of heat transfer is from warm to cold environments, not the other way around. These critiques could be drawn out from Savannah, Dev, and Ava's critique of Jasmine and Ethan's model. For example, in response to Jasmine and Ethan's explanation of their model, one PST asked, "Do you think there are different kinds of particles?" to help them think further about what the particles were representing.

Simulation Specialist Training

For simulation specialists to use the Mursion® system, they must receive Mursion®-supported training and pass a final "checkout" assessment with a Mursion® trainer. Initial training for a particular classroom (e.g., upper elementary school or middle school) involves about 60 hours of training—30 hours of synchronous training with a Mursion® trainer and 30 hours of asynchronous training. After this initial training, subsequent Mursion® training—about 15 hours, mostly asynchronous—must be added for simulation specialists to learn to enact different classrooms. Mursion training aims to help simulation specialists learn the avatars' personalities and backgrounds (e.g., number of siblings) and become proficient at selecting the avatar's facial expressions and body movements—either one at a time or in concert—using a game controller and Mursion® software. The specialists learn how to alter their own voices, use additional voice modulating software in Mursion® when needed, and employ vocal signatures to make the avatars auditorily unique. Simulation specialists learn to have discussions across the students, use various gestures, respond in character, and respond to the requests of the teacher during multiple practice sessions.

Project training, which follows Mursion® training, involves roughly 15 more hours of largely synchronous work. For the Keep It Cold discussion, there are six training lessons to help simulation specialists understand and represent students' ideas and the ways that the students can learn during the discussion. During these lessons, the simulation specialist learns about the task, the prior activities the students participated in, the models each group created, their ideas about heat transfer in the two cups of cold water in the Keep It Cold investigation, and what would need to happen during the discussion for each group to change their original thinking. The simulation specialist works with a trainer, who is a science content and teaching expert, to practice the student avatars' responses about the Keep It Cold investigation and prior activities and to rehearse full 20-minute discussions using varying teacher approaches. Trainers provide feedback to simulation specialists about where they are strong and where they need to improve, offering additional training as necessary.

METHODS

Participant Sample

We sent a call to recruit secondary science PSTs to teacher educators at our project's partner institutions. Those teacher educators then shared the recruitment flyer with their PSTs, mostly through email. In the materials, we advertised for a paid $200 opportunity asking participants to complete two surveys, prepare for and complete the simulated discussion, and complete an interview after the discussion. In the final sample, we tried to get as much variability as we could, but low response rates combined with cancelations were limiting in this respect. Based on PST availability and responsiveness, we scheduled sessions with eight PSTs.

The eight PSTs identified their gender, race, and ethnicity as follows: six female and two male; seven White and one Black; and one Hispanic/Latino and seven not Hispanic/Latino (see Table 1). Four PSTs recently completed a bachelor's degree. All PSTs had completed a science methods course and at least one science content course designed for K-12 teachers. All PSTs indicated that they had experience participating in science discussions and had studied their importance, but only half (*n*=4) indicated they had some or a little experience leading science discussions. None had prior experience using simulated classrooms. Throughout this paper, we identify the PSTs by ID (e.g., PST A).

Table 1

Participating Preservice Teachers' Demographic Data (n=8)

Demographic Data							
PST ID	Gender	Race	Bachelor's degree	Completed at least one science content and science methods course	Participated in science discussions	Led science discussions	Prior simulated classroom experience
PST A	F	White	No	Yes	Yes	No	No
PST B	M	White	Yes	Yes	Yes	Yes	No
PST C	F	Hispanic/ Latino & White	No	Yes	Yes	Yes	No
PST D	F	Black/ African American	No	Yes	Yes	No	No
PST E	F	White	Yes	Yes	Yes	Yes	No
PST F	M	White	Yes	Yes	Yes	No	No
PST G	F	White	Yes	Yes	Yes	Yes	No
PST H	F	White	No	Yes	Yes	No	No

Seven different teacher preparation programs were represented across this sample of eight PSTs. One PST was enrolled in a master's degree program, two PSTs were enrolled in five-year bachelor's degree programs, and the remaining PSTs (*n*=5) were enrolled in bachelor's degree programs. Over half of the PSTs were majoring in secondary education (*n*=5), with four of them dual majoring in natural sciences and secondary education. All eight PSTs were pursuing certification at the secondary level (including grades ranging from: 5-12, 6-12, or 7-12) and reported that they had taken more than six college-level science courses, including science courses that focused on pedagogy. Three PSTs had previous experience working in schools, two as substitute teachers and one as a tutor.

Data Collection

We gathered data from four complementary data sources for each PST: a background survey, an avatar-based simulation performance, a post-session survey, and a semi-structured interview. Prior to facilitating the science discussion in the simulated classroom, each PST completed an online background survey to provide information about their personal and professional characteristics and experiences, including their current major, previous teaching experience, previous science content and pedagogy courses taken, and previous experience facilitating science discussions and using simulated classrooms. The avatar-based simulation performances took place over Zoom between the simulation specialist and the PST; all were video recorded. Each avatar-based simulation session began with a non-content specific warm-up activity for PSTs to become familiar with the simulated classroom and student avatars. This warm-up was followed by the science discussion, which was at maximum 20 minutes. Immediately following the discussion, each PST completed a survey asking them about their discussion preparation, their perceptions on the written task, and how well they felt they facilitated the discussion. Then each PST participated in an interview that focused on their perceptions of the clarity and usefulness of the task materials, the authenticity of the task, and the importance of argumentation, as well as their thoughts on the value of using this type of performance task in teacher preparation. All discussion performances and interviews were transcribed for later analysis.

Within the survey, PSTs responded to Likert, closed-ended, and open-ended questions. Questions included asking how successful they felt they were in their science discussion session, whether they thought the amount of time they spent preparing was sufficient, whether the materials they were sent to prepare for the session were clear, and what the goals for the discussion were. PSTs also answered questions about the simulated environment, including whether: the student avatar responses were typical for middle grade students, the PSTs felt their performance accurately reflected their teaching abilities, and the PSTs felt that these kinds of simulations would be appropriate to include in a teacher preparation program. In the interview, we asked similar questions about the goals of the discussion, but also asked about the clarity and usefulness of each of the task components, as well as if any challenges with the materials arose when preparing for the discussion. We also asked about the most and least helpful parts of the materials, what they would have done differently if they could do it again, and the usefulness of the discussion task for teacher preparation.

Data Analysis

Our research team used a convergent parallel mixed methods approach to answer the study’s research questions (Creswell & Plano Clark, 2011). We used descriptive statistics to summarize the PSTs’ responses on each of the Likert scale or close ended survey items to understand their perceptions about the performance task and simulated classroom. For each set of responses to Likert scale questions, we calculated the number and percentage of PSTs who provided specific answers and noted patterns across the responses. We conducted an iterative qualitative analysis of PSTs’ responses to open-ended survey items and interview questions (Creswell, 2009; Maxwell, 2013). We applied one or more codes to each response, some of which we anticipated from previous research (Mikeska & Howell, 2021) and some that emerged from our analysis.

We analyzed the PSTs’ discussion approaches by reading through the discussion transcripts and determining whether there was evidence of the PST attending to three specific content features (direction of heat transfer, speed of particles, and differences in heat transfer between two cups) and engaging in three pedagogical features (encouraging use of evidence to justify ideas, encouraging critique, and referencing the learning goal) during the discussion. We identified these six features and used them in the analysis, as they were the ones that the task was intended to prompt the PSTs to take up during the discussion. We also analyzed where the PSTs decided to start their discussion: either with the student avatars’ models or by asking the student avatars to think back to the previous investigations.

We scored each discussion using a previously developed four-level scoring rubric to measure teaching performance across five key dimensions of high-quality argumentation-focused discussions (Mikeska et al., 2019). The four levels of the rubric are: (1) beginning practice, (2) developing practice, (3) well-prepared practice, and (4) commendable practice. The five key dimensions of high-quality argumentation focused discussions are: attending to student ideas (dimension 1), developing a coherent storyline (dimension 2), encouraging student-to-student interactions (dimension 3), developing students' conceptual understanding (dimension 4), and engaging students in argumentation (dimension 5).

For this study, three raters used the rubric to score each discussion. To prepare for scoring, the raters completed training using a series of online webinars and related documents to develop a shared understanding of the rubric and how to apply it consistently across discussions. First, each rater read through the scoring rubric document; each dimension was described in-depth and had two or three related indicators to explain its focus. For example, dimension four, which focused on the extent to which the PST adequately developed students' conceptual understanding, was comprised of three indicators. One indicator assessed the extent to which the PST provided opportunities for students to evaluate one another's ideas (versus the teacher being the one engaged in the evaluation of student ideas). Another indicator evaluated the extent to which the PST made any incorrect or imprecise statements about the content during the discussion. The final indicator evaluated how well the PST addressed key student misunderstandings during the discussion – ideally by having other students critique, offer rebuttals, and use evidence-based reasoning to persuade each other. Each indicator included specific observable characteristics at each of the four scoring levels; detailed observer notes also accompanied each scoring dimension to help raters know how to make decisions between scoring levels. Second, each rater completed a series of seven different webinars – one for each scoring dimension; one to describe the overall rating process and logistics; and one about how to be aware of and address bias during scoring. The five dimension-specific webinars explained the indicators for each dimension, reviewed the different scoring levels, and provided the raters with opportunities to practice scoring video clips on these dimensions and indicators. Finally, each rater completed and received feedback on their scoring for one full practice video; they also met with a scoring training lead to review their scores prior to beginning the scoring of the study videos. Exact initial rater agreement was 70% across dimensions for two of the eight videos, with disagreements reconciled.

After the raters scored the discussions, we analyzed the scored discussions to examine whether this performance task elicited adequate variability in PSTs' ability to engage in this teaching practice. This analysis involved comparing the number and percentage of PSTs who scored at each scoring level within and across the five scoring dimensions, as well as examining each PST's score profile across the five dimensions to determine if their scores were consistent or varied across dimensions.

FINDINGS

RQ1: PSTs' Perceptions of the Performance Task and Discussion Implementation

Participants shared their perceptions of the clarity and helpfulness of seven different sections of the task materials (e.g., Introduction, Lesson Overview, etc.) in preparation for the discussion. All PSTs rated the clarity of the materials to be somewhat or very easy to understand in all sections, except for one PST who did not rate the video examples section because they did not review it. All PSTs reported that the Introduction to the Keep It Cold task, which introduces the learning goal and describes the goal of the discussion from the PST's perspective, and the Teaching Tips, which are scattered throughout the task with reminders about important ideas to attend to and how to most efficiently interact with the students, were very easy to understand. The remaining sections received a mix of very or somewhat easy to understand ratings. PSTs also reported that all sections were somewhat or very helpful in preparing for the discussion. The majority of PSTs (*n*=7) reported that the Teaching Tips section and the Student Responses/ Making Sense section, which provides the students' written work and highlights their understandings and misconceptions, were the most helpful sections.

In terms of their understanding of the task goal, all PSTs reported that the primary discussion purpose was to reach a consensus about the most effective model of heat transfer. For example, one PST indicated that the goal was "to complete a consensus model by using critiques of past models [those already made by the students] and to use evidence in the creation of a representation." Another PST mentioned how the task's goal was to "facilitate a discussion about heat transfer…and getting these two groups of students to develop a consensus model of heat transfer based on a specific example of water at different temperature." Most PSTs also noted secondary purposes for the discussion, such as focusing on argumentation and making connections to previous work (*n*=6), encouraging student to student discourse (*n*=4), correcting student misconceptions (*n*=3), making meaning from the experiment (*n*=1), speaking and thinking like scientists (*n*=1), and leading students to the right answer (*n*=1). For example, one PST responded that their goal was "to have all students participate equally and get students to correct one another's misconceptions." Another PST explained how they needed to ensure that the students "discuss their ideas with each other," which highlights the goal of encouraging direct student discourse, while a different PST commented on making sure they developed and used questions that "get the students to…provide evidence….and come to a consensus" about the model, which supports the focus on argumentation.

When asked to rate their overall success in meeting the discussion's student learning goal, most PSTs responded that they had been somewhat (*n*=6) or very successful (*n*=2). Areas in which all PSTs reported that they were very or somewhat successful included incorporating key ideas in students' written prework and facilitating a discussion that is organized, purposeful, and focused on the content at hand. However, PSTs felt least successful in promoting student interaction and making precise statements about the science content to help students work towards correct understandings. All PSTs reported that if they were given more opportunities to practice in the simulated classroom their performance would improve.

Overall, the PSTs described the task as reasonable for middle school, aligned with their past experiences with students at this age, and that it was appropriate in terms of middle school content. For example, one PST noted how even though "combining like heat transfer and particle movement might be a little challenging…it's something I certainly would do with middle school students." Another PST made a direct connection to their student teaching by stating that, "…it seemed like the class I was in for student teaching, like we did that kind of stuff all the time, so it was de nitely a familiar process." Other PSTs noted how in this discussion the students were "able to use evidence from their experiment to have thoughts about how the science concept work[s]," which was similar to what they would typically do as a middle school teacher.

When asked how typical the responses and behaviors of the student avatars during the discussion were compared with the responses and behaviors of actual students at this grade level, all PSTs reported that the avatars behaviors were very typical (*n*=2) or somewhat typical (*n*=6). For example, one PST noted how "it took me off guard how natural it was…it owed really well, it made sense…it was unnatural how natural it was" while another PST stated that it "felt like a real discussion." However, one PST explained how "they were able to like overturn their misconception really quickly, which usually students hold on [to] misconceptions a little bit longer." Some PSTs also noted dif culty seeing the avatars' non-verbal expressions and that no access to a whiteboard contributed to reduced authenticity.

When asked how appropriate or important they believe a discussion task like this would be as a component of a teacher preparation methods course, the PSTs thought it would be very (*n*=6) or somewhat important (*n*=2). All PSTs agreed that it was an experience that should be included in teacher preparation due to the limited in-class practice teaching experiences currently provided and the ways in which this task gave them an opportunity to facilitate a discussion without having to write a full lesson plan. One PST explained how such an experience provides teachers with "experience with facilitating a discussion or develop[ing] a plan for facilitating a discussion or how to practice argumentation" which serves as "good practice for students before they actually get in front of a real classroom." Another PST explained how providing these learning opportunities in teacher education courses would be similar to giving them "training wheels" where they could practice putting all the pieces together before having to engage real students in these kinds of science discussions.

RQ2: PSTs' Discussion Performances and Approaches to Facilitating Argumentation-Focused Discussions

The PSTs demonstrated variability in the quality of the discussions they facilitated in the simulated classroom and in the approaches they used. We begin this section by providing a broader picture of the quality of the discussions these PSTs facilitated related to ve dimensions of this practice. We end by discussing the extent to which they addressed key content and pedagogical features noted in the designed task during the discussion, as well as provide an in-depth look at these similarities and differences in approaches across two PSTs.

PSTs' Discussion Performances

Using the results from rater scoring, we observed similarities and differences in how well the PSTs were able to address the ve key dimensions of high-quality discussions. As shown in Table 2 and Figure 4, these PSTs were most successful at attending to students' ideas, with moderate success in facilitating a coherent discussion, developing students' conceptual understanding, and engaging students in argumentation. In comparison, these PSTs were less successful at prompting direct student interaction. As shown in Figure 5, ndings also showed variability across the PSTs' dimension scores and their overall scores (range of 8 to 18 total points out of 20), illustrating how PSTs could be stronger in certain areas and their discussion performance varied across these ve dimensions of this teaching practice. This gure shows how there were six unique scoring pro les (two pairs of PSTs shared the same scoring pro le) across these ve dimensions of practice. All PSTs' dimension scores varied across two or three different scoring levels.

Table 2
Descriptive Statistics for Preservice Teachers' Discussion Scores by Scoring Dimension (n=8)

Dimension	Mean	Standard Deviation	Range	Median
Attending to Student Ideas (Dimension 1)	3.25	0.89	2.00	3.50
Facilitating a Coherent and Connected Discussion (Dimension 2)	3.00	1.07	3.00	3.00
Encouraging Student-to-Student Interactions (Dimension 3)	1.88	0.64	2.00	2.00
Developing Students' Conceptual Understanding (Dimension 4)	3.13	0.64	2.00	3.00
Engaging Students in Argumentation (Dimension 5)	3.00	0.76	2.00	3.00
Total Score	14.25	3.62	10.00	15.00

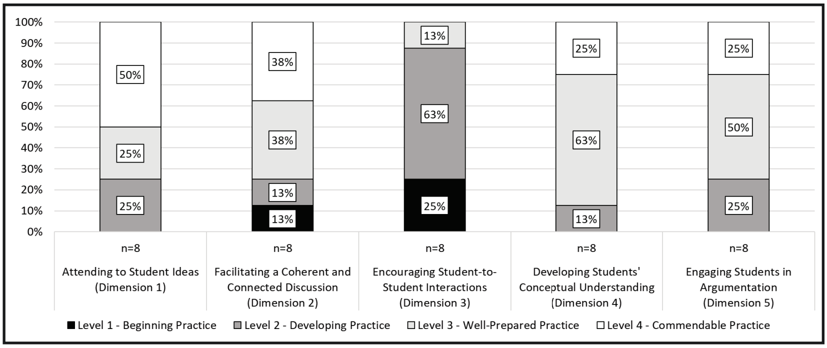

Figure 4. Preservice Teachers' Discussion Scores by Scoring Dimension (n=8).

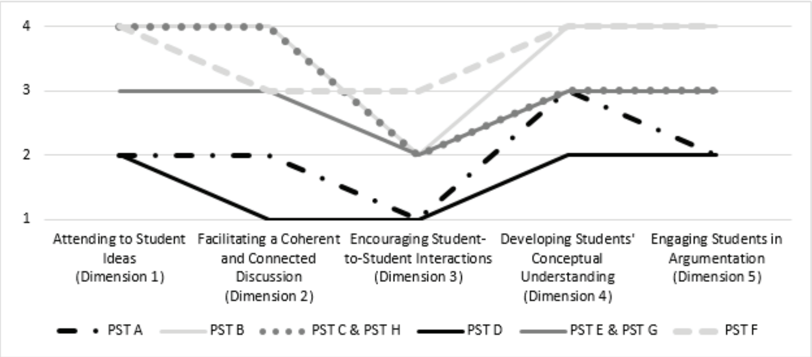

Figure 5. Preservice Teachers' Discussion Scores across Scoring Dimensions (n=8).

PSTs’ Discussion Approaches

Table 3 provides a summary across all PSTs of the three content and three pedagogical features, as described in the Data Analysis section, that were evident (or not) in the discussions. PSTs largely addressed the content features that they wanted the students to include in their models of heat transfer: arrows indicating the direction of heat transfer (6 PSTs), the speed of particles (8 PSTs) and differences in heat transfer across the foam and paper cups (8 PSTs).

Table 3

Comparison of Content and Pedagogical Features in Preservice Teachers’ (PSTs) Discussions

	Discussion Features					
PST ID	Attends to arrows indicating direction of heat transfer	Attends to speed of particles	Attends to differences in heat transfer across foam and paper cups	Encourages use of data table as evidence to justify or refine ideas	Encourages students to engage in critique of the other group’s model and/or ideas	References the learning goal during the discussion
PST A	No	Yes	Yes	Yes	No	Yes
PST B	Yes	Yes	Yes	Yes	Yes	Yes
PST C	Yes	Yes	Yes	No	Yes	No
PST D	Yes	Yes	Yes	Yes	No	No
PST E	Yes	Yes	Yes	Yes	Yes	Yes
PST F	Yes	Yes	Yes	Yes	Yes	Yes
PST G	Yes	Yes	Yes	Yes	No	Yes
PST H	No	Yes	Yes	Yes	Yes	Yes
Total	75% (6)	100% (8)	100% (8)	82.5% (7)	62.5% (5)	75% (6)

Note. “Yes” means that there was evidence of that content or pedagogical feature within the science discussion the PST facilitated, while “No” means that feature was not present within the science discussion the PST facilitated in the simulated classroom.

Despite this consistency in addressing these three content features, the PSTs took unique paths in their discussions. Six PSTs began by having the students reference the models that they created. Five of those six PSTs asked students to say what they agreed with or disagreed with in the other group’s model. The prior investigations were the starting point for the remaining two PSTs. One asked the students to look specifically at the Heat Conduction activity to start from a common understanding of heat transfer before they discussed how they modeled it. The other PST who started by asking students about what all the previous investigations had in common reported that this was to reach consensus on the direction of heat transfer before discussing the models. Those two PSTs then asked students about the features of their models and what should be kept in the final model.

All PSTs worked throughout the discussion to at least modify the students' existing models of heat transfer, with five PSTs successfully addressing the learning goal to create a single consensus model. The remaining three PSTs either had students modify their existing models or talked broadly about what would be included in a final model; however, they did not have the students actually create the new consensus model during the discussion.

In addition to asking students what they agreed or disagreed with in each other's models (five PSTs), seven PSTs asked the students to explain their agreement or disagreement by citing evidence from the previous investigations to accompany claims about the direction of heat transfer, the motion of particles, and differences in heat transfer between the foam and paper cups. While levels of agreement is a simple form of argument critique, the fact that most PSTs pressed the students to cite evidence from the previous investigations highlights the potential that tasks like this, specifically designed for argumentation, have to support PSTs to learn to engage in this ambitious teaching practice. For example, one PST said to the student avatars, "Hearing Ethan's statement and taking a look at the data, Savannah, Dev and Ava, what do you think about modeling heat transfer with the foam cup?" This PST was making use of both Ethan's critique and the data from the previous investigation to encourage Savannah, Dev and Ava's group to reconsider their model with this information in mind.

To illustrate some further variability in approaches and performance in the discussions, two PSTs' discussions, that of PST A and PST B, are summarized here and depicted in Figure 6. They were chosen because they started their discussion in different ways and their discussion scores differed quite a bit with PST B scoring much higher across the dimensions of high-quality argumentation-focused discussions than PST A.

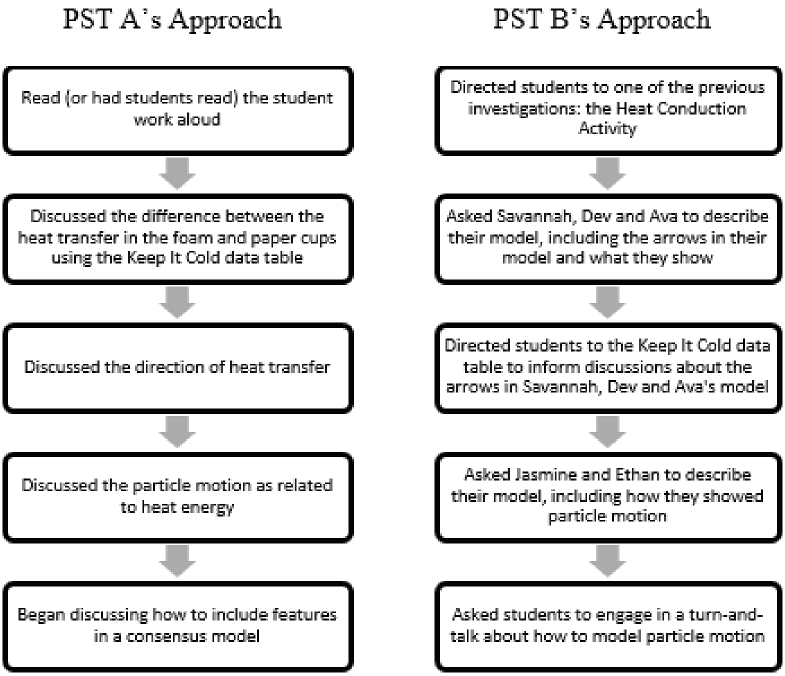

Figure 6. Comparison of PST A (lower scoring) and PST B’s (higher scoring) Discussions.

PST A. PST A started the discussion by focusing on Savannah, Dev and Ava’s model. After reading their full response as it appeared in the task packet aloud to the class, PST A asked Jasmine and Ethan to read what they wrote about their own model aloud. This PST was the only one to spend time reading the students’ responses aloud, even though the students had already reviewed each other’s models and wrote critiques of them. PST A then encouraged the students to look at the data table from the Keep It Cold investigation to notice that, contrary to their model, heat was still entering the foam cup because the temperature of the water in it was increasing. The students decided that their arrows should not stop at the edge of the foam cup, but should still enter the cup, just less so than in the paper cup. Again calling on prior investigations to support the discussion, PST A brought up the Heat Conduction activity involving pats of butter, a candle and ice to help them remember that they learned that heat moves from warmer to cooler areas.

After settling the issue of the direction of heat transfer, the class turned their attention to the particles. By pressing the students to think about a continuum of cold and warm particles, PST A encouraged the students to talk about particles with respect to the speed at which they were moving instead of the kind of particle they were. The class then began talking about creating a consensus model with time running out. The PST asked the class, “What can we agree on as far as the heat transfer and maybe the reason why the foam cup and the paper cup are different?” Dev pointed out that heat enters both cups but enters the foam cup more slowly than it does the paper cup. The PST recorded this idea on a piece of paper with thick (the paper cup) and thin (the foam cup) arrows entering the cup and showed it to the class. PST A then asked the students, “What would the particles look like?” At that point, time ran out and the session ended.

In summary, PST A read the student work to the class, addressed the difference in insulation properties of the foam and paper cup, discussed the direction of heat transfer and then discussed the particles as moving at different speeds instead of being different kinds of particles. Then the PST focused the class’s attention on the consensus model and asked them what they could agree on to include in it. Although PST A addressed the key misunderstandings that were evident in the students’ models, this PST struggled to engage the student avatars in extensive argumentation or interaction with one another during the discussion.

PST B. Overall, PST B demonstrated more skill in facilitating an argumentation-focused discussion among the students. PST B started his session by asking the students to look back at the Heat Conduction activity with the candle, pats of butter and ice. The PST asked the class, “Now, from that activity, what did you notice was happening?” PST B pressed the students to consider the direction of heat transfer by rst asking which pats of butter were melting rst. PST B went a step further to ask the students, “What type of energy is that candle producing?” This was followed by asking which direction that heat energy was moving and encouraged the class to reach consensus on the direction of heat transfer. PST B said, “Savannah, Dev, Jasmine, and Ethan, what do you think about Savannah’s idea that heat travels from a hot to a cold?” Much like PST A, PST B was using the prior investigation to help the students focus on the direction of heat transfer before thinking more speci cally about the features that were in the students’ models. However, PST B went a step further by having them consider each other’s ideas and respond to them directly to build towards consensus on this point.

To make sure the class agreed on the direction of heat transfer, and keeping the discussion close to the models themselves, PST B asked Savannah, Dev and Ava to describe their model. After they did, PST B asked Jasmine and Ethan, “Now that we talked about the butter activity, I know you didn’t agree with the direction of their arrows. How do you feel about the direction of the arrows showing heat transfer in their model?” Once Jasmine and Ethan both agreed on the direction of heat transfer, the PST said, “If you have questions about the idea of the movement of heat, we can revisit that...For right now let’s work with the common definition that heat transfers...from hot to cold.” With these talk moves, PST B was linking the discussion to the students’ work and checking in to be sure that the whole class agreed on the takeaway from the Heat Conduction activity. In these ways, PST B was working toward a coherent and connected discussion.

Similar to PST A’s session, PST B encouraged the students to look at the Keep It Cold data table for evidence. PST B did not name the data table like PST A did, though. Rather, PST B asked Savannah, Dev and Ava, “Is there any evidence or any data that can support your model…showing that the heat was increasing in the cup?” The students themselves then referred to the data table and said, “If you look at the data table, the temperature of the water increased” in both cups. Once the students reached agreement around the fact that heat was entering the foam cup, PST B paused the discussion to remind students of their end goal: to create a consensus model to explain heat transfer. PST B said, “How can we show that these cups are different and that heat transfer is different because these cups are made of different materials?” So, in comparison to PST A, PST B was addressing the key misunderstandings represented in the models, but was also supporting the students during the discussion to keep in mind why they were talking about those ideas: to ultimately create one model that explained what was happening in the two cups. In this sense, PST B was keeping the discussion connected to the learning goal while also building off the students’ contributions.

A feature that PST B included in their discussion that was missing from PST A’s discussion was engaging students in conversation with one another. PST B asked the students to engage in a turn-and-talk with one another to decide how they could represent what was happening in the two cups in their model. Throughout, PST B also made sure all students’ voices were heard. Additionally, PST B made sure there was consensus among the five students before moving on to a new idea.

Much like PST A, PST B also discussed the particles and addressed the fact that the particles were not different but were just moving at different speeds. PST B had the students talk to one another again to decide how to represent movement of the particles in the cups. While PST B also ran out of time, PST B concluded the discussion by saying, “We’re going to see if we can’t bring both of your ideas, the arrows in the heat transfer and the particle movement together into one consensus model to really represent, in a model, what’s happening with heat transfer.” Here, PST B once again attributed the ideas to the students and reminded them of the purpose of their discussion: to create a consensus model.

DISCUSSION

Previous research has tended to focus on approaches and strategies for supporting students’ engagement in scientific argumentation and examining various factors that support or hinder students’ productive engagement in this important science practice (Colley & Windschitl, 2016; McNeill, 2011; McNeill et al., 2016, 2017). One key factor points to the importance of the teacher’s role and their beliefs, understanding, and teaching skills in this area (Berland, 2011; Driver et al., 2000; Knight-Bardsley & McNeill, 2016). That is, research has suggested that students’ productive engagement in scientific argumentation is directly related to how science teachers set up and support that engagement within classroom interactions and activities. More recent research has included studies that use specific tools and approaches to help science teachers learn how to support students to engage in scientific argumentation, especially as part of whole class and small group discussions (e.g., Marco-Bujosa et al., 2017; Mikeska et al., in press; Osborne et al., 2019). Our findings suggest that the use of a performance task set in the context of a simulated classroom has potential as an innovative tool that secondary science teacher educators can leverage to help PSTs learn to engage in this teaching practice.

First, study findings indicate these PSTs saw value in this performance task as a useful tool that could be integrated into secondary science method courses to provide a practice space for them to improve in their ability to facilitate these kinds of discussions. This is consistent with other studies that have shown promise in using performance tasks within simulated classroom environments to help PSTs engage in ambitious teaching practices (Bell, 2019; Mikeska & Howell, 2020; Straub et al., 2015). There was also strong agreement that the task authentically represented the work that secondary science teachers engage in when facilitating discussions and the student avatars responded in ways typical of middle school students. These findings are similar to other studies that examined PSTs’ perceptions of the authenticity and usefulness of such performance tasks, albeit within elementary mathematics and science methods courses (Mikeska & Howell, 2021).

Second, findings provided empirical evidence that the PSTs understood the primary goal of the science discussion they were being asked to facilitate in the simulated classroom. PSTs' similar conceptions of the discussion's primary student learning goal likely was a reason that there were similarities in the content and pedagogical aspects they addressed within their discussions. We observed that there was strong attention to the three key aspects of the consensus model noted in the performance task as important to address; all PSTs provided opportunities for the students to consider how to illustrate the speed of the particles and the difference in heat transfer with respect to the foam and paper cups. Similarly, most PSTs attempted to use key teaching moves, such as encouraging the use of data to justify or refine ideas and encouraging students to critique others' ideas, noted in the performance task as important for supporting students' productive engagement in the argumentation-focused discussion. These argumentation-specific moves have been shown to support students in key aspects of argument construction and critique (Gonzalez-Howard & McNeill, 2019; Grooms et al., 2018; Jimenez-Alexandre & Erduan, 2008; Mikeska & Howell, 2020). Overall, the PSTs understood the task purpose and the various content and pedagogical features they needed to attend to when facilitating the discussion. Despite this shared understanding, however, there was variability in how well the PSTs were able to engage in certain dimensions of this practice.

These PSTs' discussion performances illustrated variability in skill within and across the five dimensions of this teaching practice, similar to variability we have observed with elementary PSTs engaging in similar tasks (Mikeska et al., in press). This variability suggested that PSTs still had room to grow within at least one but typically within multiple dimensions to reach the "commendable" scoring level. This is in keeping with the notion that facilitating argumentation discussions is challenging for teachers to learn how to do well (e.g., Osborne et al., 2013). In addition, no PST scored at the same level across all five dimensions. Instead, their individual scoring profiles indicated that their discussion facilitation skills had both strengths and areas for growth. This finding points to the potential viability of such tasks as tools to engage PSTs in productive approximations of practice, as their scores did not illustrate a ceiling effect and left room for improvement over time.

Across the study participants, results showed that the PSTs encountered the most difficulty in the third scoring dimension – encouraging student to student interactions – during the discussion. This finding aligns with previous research showing that it can be difficult for teachers to facilitate student led discussion where students are given opportunities to interact directly with each other to support scientific sense-making (Davis et al., 2006). Instead, research has shown that teachers tend to engage students in interaction patterns where the teacher controls the discussion turn-taking and evaluates student responses – what has been coined the initiate-respond-evaluate (IRE) discourse pattern (Cazden, 1988; Lemke, 1990). This finding suggests that this aspect might be a good starting place for teacher educators to provide additional scaffolding for PSTs to improve in this teaching practice.

LIMITATIONS

There are three key limitations in this study. The first of these is that we used a sample of convenience; this sample is not representative across all secondary science PSTs in the United States, which limits the generalizability of the study's findings. In previous research, we found that working with approximately five to eight PSTs provides sufficient variation to examine their perceptions of the task's clarity, usefulness, and authenticity and their approaches to facilitating the specific science discussion for one performance task (Mikeska & Howell, 2020).

The second limitation is that we used self-report data from study participants. This comes with the possibility of response bias where participants share responses that they think the researcher wants to hear instead of their actual perceptions. To decrease the possibility that participants would only provide positive feedback, we clearly communicated to each PST that the goal of this study was to better understand how the performance task was functioning and ways that it could be improved to better support PST learning in the future when we integrate the use of the performance task into teacher education programs (in our case, secondary science methods courses). We also assured them that their responses would be used to identify patterns across participants to inform future task revisions.

Finally, this study examined the PSTs' perceptions and use of the simulated classroom as they engaged in only one ambitious science teaching practice: facilitating argumentation-focused discussions. In addition, this study only focused on the use of one science performance task, which was limited to one science content area (physical science) and topic (heat transfer). It is possible that the PSTs' perceptions, performances, and approaches may have been different if the performance task had addressed other science content areas or topics. Future research could examine how these aspects might vary across content areas and topics at the secondary level.

IMPLICATIONS

The findings from this study have important implications for leveraging these types of innovative tools to support science teacher learning. Study results made clear that these PSTs saw the value in using this performance task within a simulated classroom environment to tryout an instructional practice that is perennially hard to learn how to do well. They also noted the importance of the task's goal and content focus and strong alignment to the work of middle school science teaching. Collectively these findings suggest the importance of and potential benefits of incorporating these types of practice-based learning opportunities into secondary science teacher education courses. Such learning opportunities provide a concrete solution for helping PSTs learn how to apply what they are learning about in their teacher education courses and doing so in an environment where they can take risks, make mistakes, and try again – all without harming any real students.

Results also suggest that teacher educators need to carefully consider ways to support PSTs so they can learn from these experiences. In this study, while most of the PSTs at least attempted to address similar content and pedagogical features during the Keep It Cold discussion, the scores suggest there was variability in how well they were able to adequately address these features to facilitate a high-quality, argumentation-focused discussion. In particular, study results suggest that one area ripe for additional support is in helping PSTs learn how to facilitate student-led discussions where students interact directly with each other as they justify and critique ideas and move towards consensus during the discussion. Some possibilities could involve teacher educators modeling how to facilitate student-led discussions themselves, having PSTs reflect on their own and others' discussions to identify productive teaching moves used during these discussion, providing formative feedback to PSTs on specific aspects of the discussion they facilitated, or giving the PSTs an opportunity to plan for and facilitate a second discussion using the same task to build off their strengths and address their areas for improvement from their first discussion. Future research could examine these varied mechanisms and the ways in which they – individually or in combination with each other -- support PST learning when using simulated teaching experiences.

CONCLUSION

Findings from this study indicated that the PSTs shared positive perceptions of this science performance task's clarity, usefulness, and authenticity. This nding suggests the performance task has strong face validity and has the potential to be a useful learning tool to support PSTs as they learn to engage in this ambitious teaching practice. Study results also indicated that the PSTs showed a strong understanding of the task's primary discussion goal, although their scores indicated that they had room to grow in their ability to facilitate such discussions. As such, this science performance task, as designed, created a situation where the PSTs could practice facilitating an argumentation-focused science discussion and potentially build towards improvement over time. Providing a standardized task where each PST must contend with similar student ideas and is trying to address the same student learning goal can provide productive fodder for science teacher educators to compare teachers' discussion performances and identify strengths and areas for targeted growth across groups of teachers. Such tools can also provide artifacts for analysis where teachers can re ect on the affordances and limitations of their own and others' teaching approaches. Future research can examine whether and how PSTs' ability to engage in this teaching practice improves when using these types of performance tasks within an online simulated classroom and how teacher educators integrate the use of such tasks within teacher education settings.

Author Note

This grant was funded by the National Science Foundation (grant #2037983). Any opinions, ndings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily re ect the views of the National Science Foundation. We are grateful for the science task team members (Liz Orlandi and Ken King) who supported the development of this secondary science task and for the preservice teachers who piloted and provided their feedback on the task.

References

Arias, A. & Davis, E. (2019, February 19-21). *Simulatedstudentinterviewsfor preserviceelementaryscienceteaching*[Paper presentation]. Simulations in Teacher Education Conference, Louisville, KY.

Arici, A., Barab, S., & Borden, R. (2016). Gaming up the practice of teacher education: Quest2teach. In L. Lin & R. Atkinson (Eds.), *EducationalTechnologies: Challenges, applications, and learning outcomes* (pp. 95–114). Nova Science Publishers, Inc.

Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497–511.

Bell, K. (2019, February 19-21). *Teaching math and science to avatars, oh my!* [Paper presentation]. Simulations in Teacher Education Conference, Louisville, KY.

Benedict-Chambers, A. (2016). Using tools to promote novice teacher noticing of science teaching practices in post-rehearsal discussions. *Teaching and Teacher Education, 59*, 28–44. https://doi.org/10.1016/j.tate.2016.05.009

Benedict-Chambers, A., & Aram, R. (2017). Tools for teacher noticing: Helping preservice teachers notice and analyze student thinking and scientific practice use. *Journal of Science Teacher Education, 28*(3), 294–318. http://dx.doi.org/10.1080/1046560X.2017.1302730

Benedict-Chambers, A., Fick, S. J., & Arias, A. M. (2020). Preservice teachers’ noticing of instances for revision during rehearsals: A comparison across three university contexts. *Journal of Science Teacher Education, 31*(4), 435–459. https://doi.org/10.1080/1046560X.2020.1715554

Berland, L. K. (2011). Explaining variation in how classroom communities adapt the practice of scientific argumentation. *Journal of the Learning Sciences*, *20*(4), 625–664. https://doi.org/10.1080/10508406.2011.591718

Berland, L. K., & Hammer, D. (2012). Framing for scientific argumentation. *Journal of Research in Science Teaching*, *49*(1), 68–94. https://doi.org/10.1002/tea.20446

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education, 93*(1), 26–55. https://doi.org/10.1002/sce.20286

Berland, L. K., & Reiser, B. J. (2011). Classroom communities’ adaptations of the practice of scientific argumentation. *Science Education*, *95*(2), 191–216. https://doi.org/10.1002/sce.20420

Bondie, R., & Dede, C. (2021). Redefining and transforming field experiences in teacher preparation through personalized mixed-reality simulations. In R. E. Ferdig & K. E. Pytash (Eds.), *What teacher educators should have learned from 2020* (pp. 229–242). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/p/219088/

Bondie, R., Mancenido, Z., & Dede, C. (2021). Interaction principles for digital puppeteering to promote teacher learning. *Journal of Research on Technology in Education, 53*(1), 107–123. https://doi.org/10.1080/15391523.2020.1823284

Cartier, J. L., Smith, M. S., Stein, M. K., & Ross, D. (2013). *Five practices for orchestrating productive task-based discussion in science.* NSTA: Corwin Press.

Cazden, C. B. (1988). *Classroom discourse: The language of teaching and learning*. Heinemann.

Christensen, R., Knezek, G., Tyler-Wood, T., & Gibson, D. (2011). simSchool: An online dynamic simulator for enhancing teacher preparation. *International Journal of Learning Technology*, *6*(2), 201–220. https://doi.org/10.1504/IJLT.2011.042649

Cohen, J., Wong, V., Krishnamachari, A., & Berlin, R. (2020). Teacher coaching in a simulated environment. *Educational Evaluation and Policy Analysis, 42*(2), 208–231. https://doi.org/10.3102/0162373720906217

Colley, C., & Windschitl, M. (2016). Rigor in elementary science students' discourse: The role of responsiveness and supportive conditions for talk. *ScienceEducation, 100*(6), 1009–1038. https://doi.org/10.1002/sce.21243

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*(3rd ed.).Sage Publications.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designingandconductingmixed methodsresearch* (2nd ed.). Sage Publications.

Davis, E. A., Kloser, M., Wells, A., Windschitl, M., Carlson, J., & Marino, J.-C. (2017). Teaching the practice of leading sense-making discussions in science: Science teacher educators using rehearsals. *Journal of Science Teacher Education, 28*(3), 275–293. http://dx.doi.org/10.1080/1046560X.2017.1302729

Davis, E. A., Petish, D., & Smithey, J. (2006). Challenges new science teachers face. *Review of Educational Research, 76*(4), 607–651. https://doi.org/10.3102/00346543076004607

Deale, D., & Pastore, R. (2014). Evaluation of simSchool: An instructional simulation for pre-service teachers. *ComputersintheSchools*, *31*(3), 197–219. https://doi-org.proxy-tu.researchport.umd.edu/10.1080/07380569.2014.932650

Dieker, L. A., Rodriguez, J. A., Lignugaris/Kraft, B., Hynes, M. C., & Hughes, C. E. (2014). The potential of simulated environments in teacher education: Current and future possibilities. *TeacherEducationandSpecialEducation*, *37*(1), 21–33.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scienti c argumentation in classrooms. *Science Education*, *84*(3), 287–312. https://doi.org/10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A

Fishman, E. J., Borko, H., Osborne, J., Gomez, F., Rafanelli, S., Reigh, E., ... & Berson, E. (2017). A practice-based professional development program to support scienti c argumentation from evidence in the elementary classroom. *JournalofScienceTeacherEducation*, *28*(3), 222–249.

Forzani, F. M. (2014). Understanding "core practices" and "practice-based" teacher education: Learning from the past. *Journal ofTeacher Education*, *65*(4), 357–368. https://doi.org/10.1177/0022487114533800

Francis, A. T., Olson, M., Weinberg, P. J., & Sterns-Pfeiffer, A. (2018). Not just for novices: The programmatic impact of practice-based teacher education. *ActioninTeacherEducation,40*(2), 119–132. https://doi.org/10.1080/01626620.2018.1424053

Gonzalez-Howard, M., & McNeill, K. L. (2019). Teachers' framing of argumentation goals: Working together to develop individual versus communal understanding. *JournalofResearchinScienceTeaching,56*, 821–844. https://doi.org/10.1002/tea.21530

Goodson, B., Caswell, L., Dynarski, M., Price, C., Litwok, D., Crowe, E., ..., & Rice, A. (2019). *Teacher preparation experiences and early teaching effectiveness: Executive summary* (NCEE 2019-4010). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Grooms, J., Sampson, V., & Enderle P. (2018). How concept familiarity and experience with scientific argumentation are related to the way groups participate in an episode of argumentation. *Journal of Research in Science Teaching, 55*, 1264–1286. https://doi.org/10.1002/tea.21451

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record, 111*(9), 2055–2100. https://doi.org/10.1177/016146810911100905

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice, 15*(2), 273–289. https://doi.org/10.1080/13540600902875340

Henderson, J. B., McNeill, K. L., González Howard, M., Close, K., & Evans, M. (2018). Key challenges and future directions for educational research on scientific argumentation. *Journal of Research in Science Teaching*, *55*(1), 5–18. https://doi.org/10.1002/tea.21412

Howell, H., Mikeska, J.N., Tierney, J., Baehr, B., & Lehman, P. (2021). *Conceptualization and development of a performance task for assessing and building elementary preservice teachers' ability to facilitate argumentation-focused discussions in mathematics: The ordering fractions task.* (Research Memorandum No. HYPERLINK "https://www.ets.org/Media/Research/pdf/RM-21-10.pdf" RM-21-10). ETS.

Jiménez-Aleixandre, M. P., & Erduran, S. (2008). *Argumentation in science education. Perspectives from classroom-based research*. Springer.

Katsh-Singer, R., McNeill, K. L., & Loper, S. (2016). Scientific argumentation for all? Comparing teacher beliefs about argumentation in high, mid, and low socioeconomic status schools. *Science Education*, *100*(3), 410–436. https://doi.org/10.1002/sce.21214

Knight-Bardsley, A., & McNeill, K. L. (2016). Teachers' pedagogical design capacity for scientific argumentation. *Science Education*, *100*(4), 645–672. https://doi.org/10.1002/sce.21222

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education, 94*(5), 810–-824. https://doi.org/10.1002/sce.20395

Lampert, M. (2009). Learning teaching in, from, and for practice: What do we mean?. *Journal of Teacher Education, 61*(1-2), 21–34. https://doi.org/10.1177/0022487109347321

Lemke, J. L. (1990). *Talking science: Language, learning, and values.* Ablex Publishing.

Lottero-Perdue, P. S., Mikeska, J. N., & Nester, M. S. (2022). Using preservice teachers' transcript coding of simulated argumentation discussions to characterize aspects of their noticing about argument construction and critique.

Contemporary Issues in Technology and Teacher Education, 22(1) ,105-139. https://citejournal.org/volume-22/issue-1-22/science/using-preservice-teachers-transcript-coding-of-simulated-argumentation-discussions-to-characterize-aspects-of-their-noticing-about-argument-construction-and-critique

Lottero-Perdue, P. S., Mikeska, J., & Orlandi, E. (2020). Development and teacher perceptions of an avatar-based performance task for elementary teachers to practice post-testing argumentation discussions in engineering design. In *Proceedings of the 2020 ASEE Virtual Annual Conference Content Access*. American Association for Engineering Education.

Marco-Bujosa, L., Gonzalez-Howard, M., McNeill, K., & Loper, S. (2017). Designing and using multimedia modules for teacher educators: Supporting teacher learning of scienti c argumentation. *Innovations in Science Teacher Education, 2* (4), 1–16.

Masters, H. (2020). Using teaching rehearsals to prepare preservice teachers for explanation-driven science instruction. *Journal of Science Teacher Education, 31*(4), 414–434. https://doi.org/10.1080/1046560X.2020.1712047

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach*. Sage.

McNeill, K. L. (2011). Elementary students' views of explanation, argumentation, and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching, 48*(7), 793–823. https://doi.org/10.1002/tea.20430

McNeill, K. L., González-Howard, M., Katsh-Singer, R., & Loper, S. (2017). Moving beyond pseudoargumentation: Teachers' enactments of an educative science curriculum focused on argumentation. *Science Education, 101*(3), 426–457. https://doi.org/10.1002/sce.21274

McNeill, K. L., Katsh-Singer, R., González-Howard, M., & Loper, S. (2016). Factors impacting teachers' argumentation instruction in their science classrooms. *International Journal of Science Education, 38* (12), 2026–2046. https://doi.org/10.1080/09500693.2016.1221547

McNeill, K. L., & Krajcik, J. (2008). Scienti c explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching, 45*(1), 53–78. https://doi.org/10.1002/tea.20201

McNeill, K. L., & Pimentel, D. (2010). Scienti c discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education, 94*(2), 203–229. https://doi.org/10.1002/sce.20364

Mikeska, J. N., & Howell, H. (2020). Simulations as practice-based spaces to support elementary science teachers in learning how to facilitate argumentation-focused science discussions. *Journal of Research in Science Teaching, 57*(9), 1356-1399. https://doi.org/10.1002/tea.21659

Mikeska, J. N. & Howell, H. (2021). Authenticity perceptions in virtual environments. *Information and Learning Sciences, 122* (7/8), 480-502. https://doi.org/10.1108/ILS-10-2020-0234

Mikeska, J. N., Howell, H., Dieker, L., & Hynes, M. (2021). Understanding the role of simulations in K-12 mathematics and science teacher education: Outcomes from a teacher education simulation conference. *Contemporary Issues in Technology and Teacher Education, 21*(3), 781-812. https://citejournal.org/volume-21/issue-3-21/general/understanding-the-role-of-simulations-in-k-12-mathematics-and-science-teacher-education-outcomes-from-a-teacher-education-simulation-conference

Mikeska, J. N., Howell, H., & Kinsey, D. (in press). Do simulated teaching experiences impact elementary preservice teachers' ability to facilitate argumentation-focused discussions in mathematics and science? *Journal of TeacherEducation.*

Mikeska, J. N., Howell, H., & Straub, C. (2019). Using performance tasks within simulated environments to assess teachers' ability to engage in coordinated, accumulated, and dynamic (CAD) competencies. *InternationalJournal ofT esting,19*(2), 128-147. https://doi.org/10.1080/15305058.2018.1551223

National Research Council. (2011). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. The National Academies Press.

Osborne, J. F., Borko, H., Fishman, E., Gomez Zaccarelli, F., Berson, E., Busch, K. C., ... & Tseng, A. (2019). Impacts of a practice-based professional development program on elementary teachers' facilitation of and student engagement with scienti c argumentation. *American Educational Research Journal*, *56*(4), 1067–1112.

Osborne, J., Henderson, B., MacPherson, A., Szu, E., Wild, A., & Yao, S. Y. (2016). The development and validation of a learning progression for argumentation. *Journal of Research in Science Teaching, 53*, 821–846. https://doi.org/10.1002/tea.21316

Osborne, J., Simon, S., Christodoulou, A., Howell Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. J*ournalofResearchinScienceTeaching, 50*(3), 315–347. https://doi.org/10.1002/tea.21073

Sadler, T. D. (2006). Promoting discourse and argumentation in science teacher education. *Journal of Science Teacher Education, 17*(4), 323–346. https://doi.org/10.1007/s10972-006-9025-4

Sampson, V., & Blanchard, M. R. (2012). Science teachers and scienti c argumentation: Trends in views and practice. *Journal of Research in Science Teaching*, *49*(9), 1122–1148. https://doi.org/10.1002/tea.21037

Simon, S., & Richardson, K. (2009). Argumentation in school science: Breaking the tradition of authoritative exposition through a pedagogy that promotes discussion and reasoning. *Argumentation, 23*(4), 469–493. https://doi.org/10.1007/s10503-009-9164-9

Straub, C. (2018). *Best in class leadership development: How virtual reality and avatars are changing the learning landscape*. Mursion, Inc. https://www.denasamuels.com/wp-content/uploads/2020/03/How-Virtual-Reality-and-Avatars-are-Changing-the-Learning-Landscape.pdf

Straub, C., Dieker, L., Hynes, M., & Hughes, C. (2014). Using virtual rehearsal in TLE TeachLivE™ mixed reality classroom simulator to determine the effects on the performance of mathematics teachers. *2014 TeachLive National Research Project: Year 1 Findings*.

Straub, C., Dieker, L., Hynes, M., & Hughes, C. (2015). Using virtual rehearsal in TLE TeachLivE™ mixed reality classroom simulator to determine the effects on the performance of science teachers: A follow-up study. *2015 TeachLivE™ National Research Project: Year 2 Findings.*

Thompson, M., Owho-Ovuakporie, K., Robinson, K., Kim, Y. J., Slama, R., & Reich, J. (2019). Teacher moments: A digital simulation for preservice teachers to approximate parent-teacher conversations. *Journal of Digital Learning in Teacher Education*, *35*(3), 144–164. https://doi.org/10.1080/21532974.2019.1587727

Wang, X., Thompson, M., Yang, K., Roy, D., Koedinger, K. R., Rose, C. P., & Reich, J. (2021). Practice-based teacher questioning strategy training with ELK: A role-playing simulation for eliciting learner knowledge. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–27.

Zeichner, K. (2012). The turn once-again toward practice-based teacher education. *Journal of Teacher Education, 63*(5), 376–382. https://doi.org/10.1177/0022487112445789